Aiming Towards a Consistent Series of Ocean Color Measurements

Ewa J. Kwiatkowska

REASoN Research Objective

A ***consistent series*** of (systematic) ocean color measurements from multi-instrument, multi-platform and multi-year observations based on accurate and uniform calibration and validation over the lifetime of the measurement.

SIMBIOS

http://simbios.gsfc.nasa.gov

Difficulties

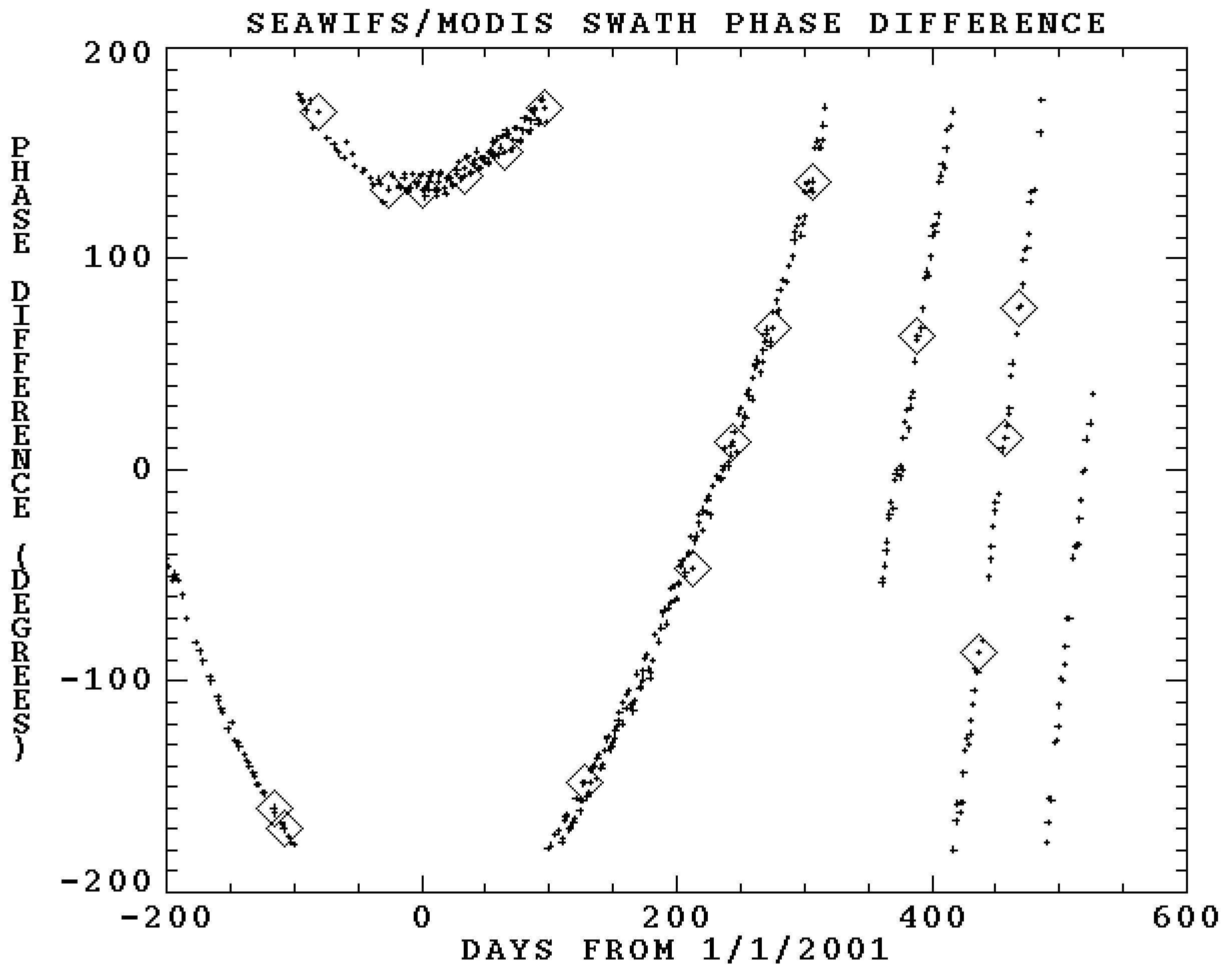
- Varying designs and characteristics of ocean color sensors.
- Disparate instrument calibrations, data processing algorithms, and validation accuracies.
- Differences in mission standard products.
- Complexity of data acquisition and retrieval.
- Management of very large data sets.

Consistent Series of Daily Global Ocean Color Data Sets

❶ Thorough multi-sensor ***data validations***:

- estimate discrepancies among data,
- extract disparate temporal trends, scan-angle dependencies, and other sensor differences,
- eliminate these trends through data cross-calibrations,
- define product accuracy levels (matchups with *in situ* measurements),
- choose the most suitable data fusion algorithms.

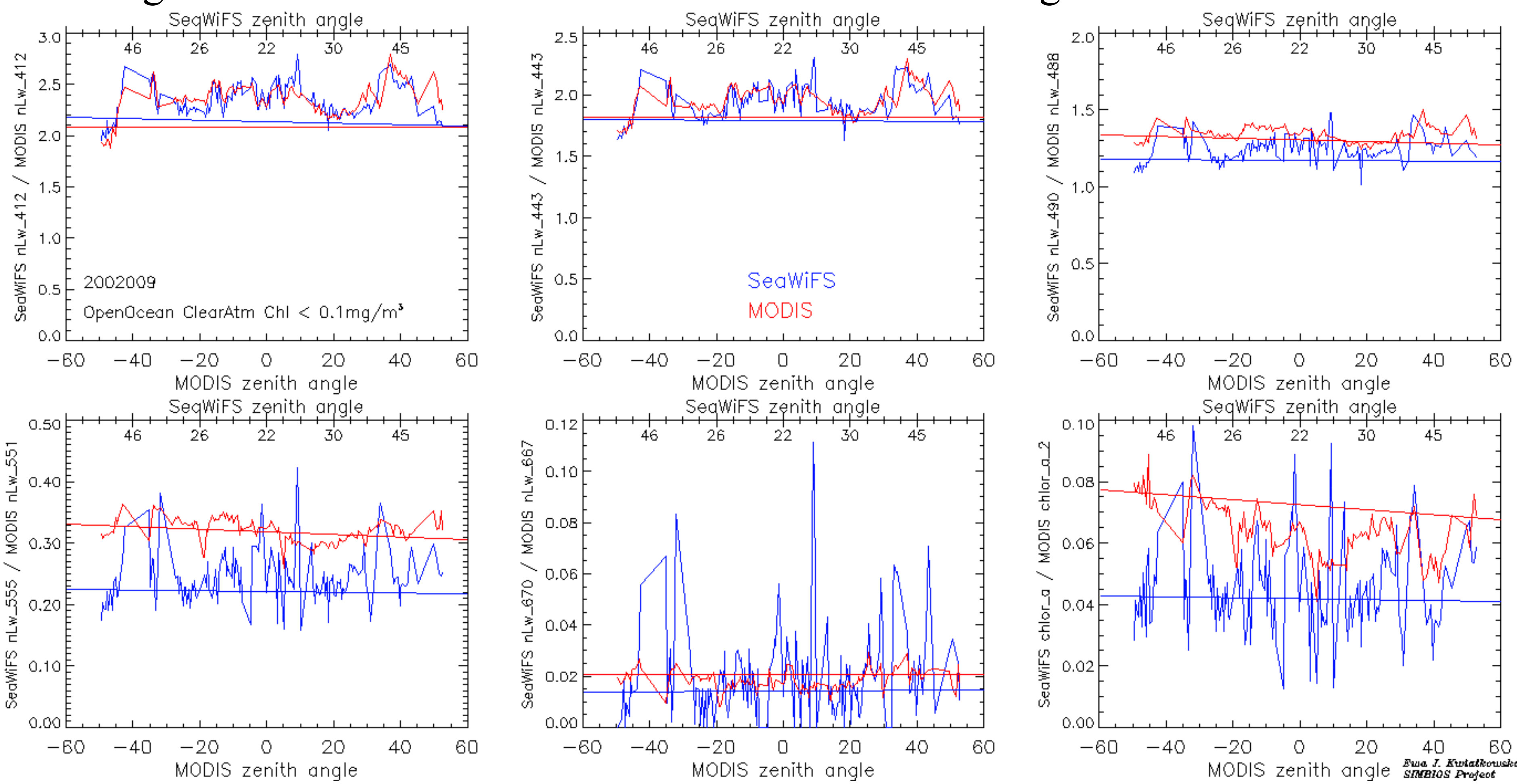
❶ Example: MODIS and SeaWiFS cross-calibration to limit scan-angle dependence in MODIS



❶ Example: MODIS and SeaWiFS cross-calibration to limit scan-angle dependence in MODIS cont.

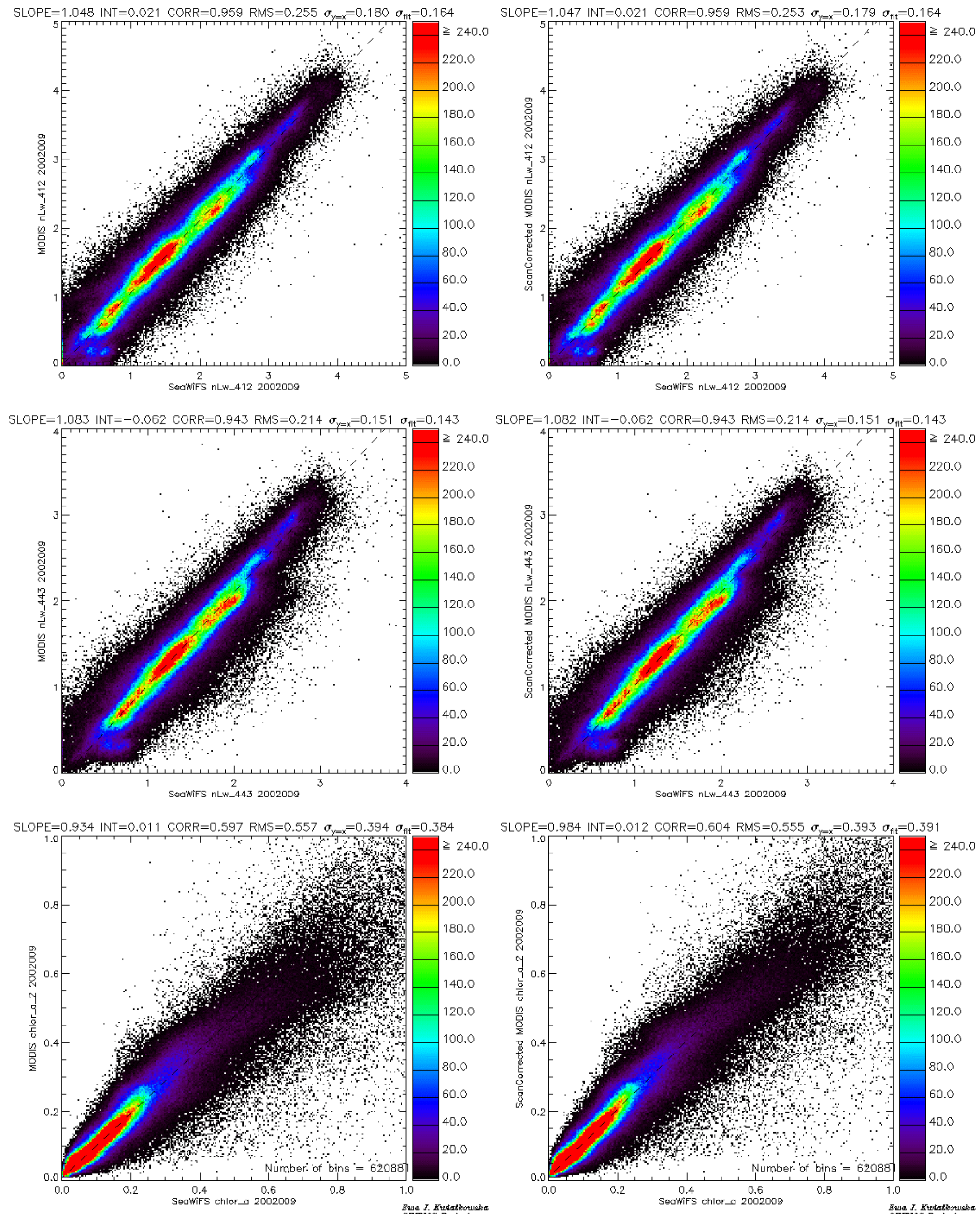
Extraction of L3b overlapping scan-line bins between MODIS and SeaWiFS

- Open ocean, clear atmosphere, and low chlorophyll data only
- 15^0-30^0 latitude ocean gyres in the Northern and Southern Hemisphere
- Search for scan lines with large amounts of overlapping data within the swath
- Average linear fit to scan-line data across sensor zenith angles over all extracted scans



❶ Example: MODIS and SeaWiFS cross-calibration to limit scan-angle dependence in MODIS cont.

Cross-calibration improved the matchups between MODIS and SeaWiFS chlorophyll based on the slope of 0.934 to the slope of 0.984.



Consistent Series of Daily Global Ocean Color Data Sets

❷ Search for suitable ***fusion approaches***:

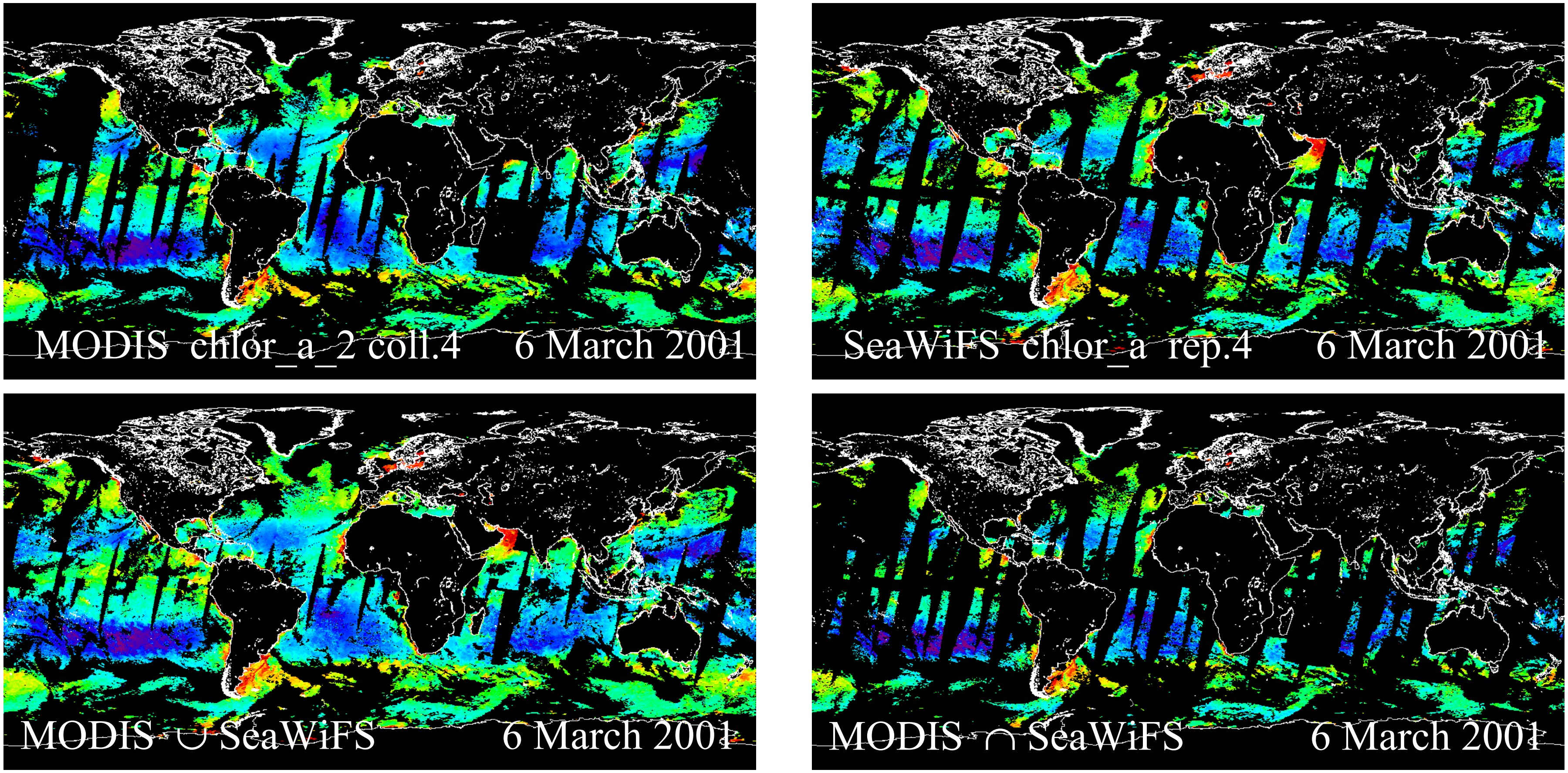
- overcome the remaining differences in sensor characteristics, instrument calibrations, data processing, and validation accuracies,
- assure a consistent accuracy through space and time for all merged data,
- assure good validation (e.g. incorporate product accuracy levels based on matchups with *in situ* measurements).

Consistent Series of Daily Global Ocean Color Data Sets

❸ ***Fusion algorithms*** currently being investigated:

- blended analysis (Watson Gregg),
- averaging and weighted averaging (SIMBIOS Project Office),
- semi-analytical optical approach (Stéphane Maritorena),
- spatial/temporal interpolation – statistical objective analysis (SIMBIOS Project Office),
- regression mapping – backpropagation neural networks and support vector machines (SIMBIOS Project Office).

❸ Example: MODIS and SeaWiFS merger



- For the purpose of data merger, averaging, weighted averaging, and semi-analytical optical algorithms are effective on overlapping sensor coverage and redundant when there is single sensor data,
- Majority of MODIS and SeaWiFS daily coverage does not overlap (~75% at 9km, although this percentage depends on the sensor swath phase difference),
- This merger produces spatially inconsistent accuracy in merged products.

Consistent Series of Daily Global Ocean Color Data Sets

4 Application of spatial/temporal interpolation and regression mapping:

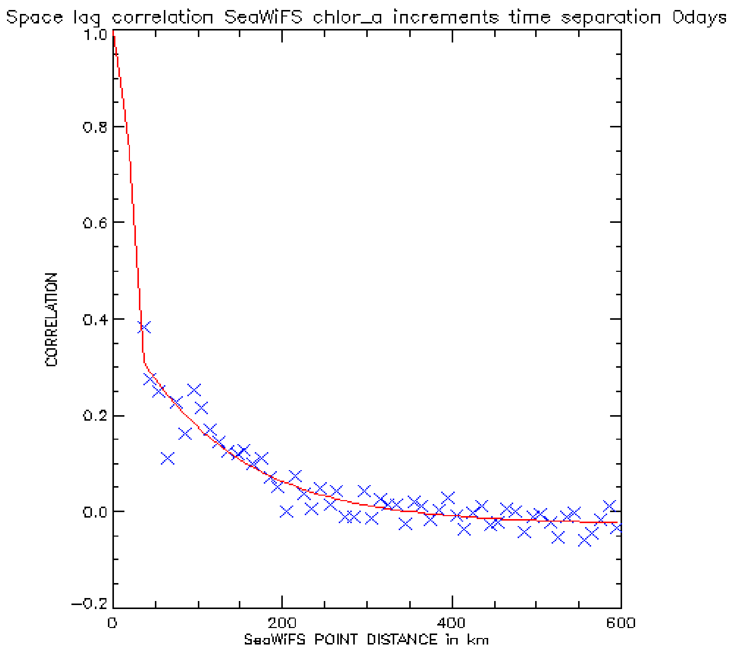
- spatial/temporal interpolation can be used to integrate ocean color products from multiple missions onto a given global grid,
- regression mapping can enable bringing multi-sensor data sets to a common ocean color baseline, consistently calibrated and jointly validated,
- spatial/temporal interpolation and regression mapping can also supplement the averaging and optical algorithms by emulating missing sensor data to provide joint sensor coverage for all merged points (this would assure consistent accuracy in merged data products, i.e. accuracy through space).

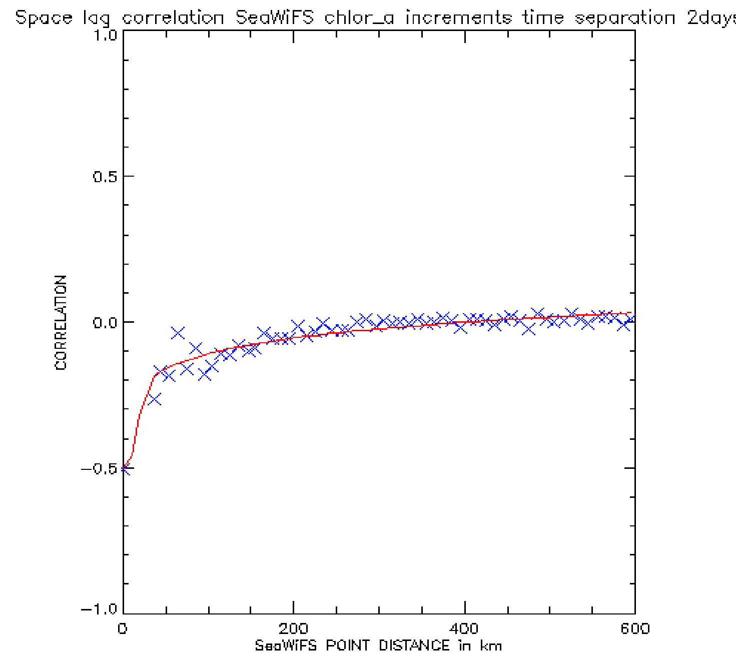
Statistical Objective Analysis
example: SeaWiFS chlorophyll

- Space (and time) interpolation of ocean color data.
- Uses chlorophyll signal-to-noise ratio.
- Dependent on the definition of the chlorophyll variance and the variance of chlorophyll noise.
- Dependent on the choice of the first-guess background field (e.g. previous day chlorophyll field, chlorophyll temporal mean, climatology).
- Uses ensemble spatial/temporal correlation structure of the chlorophyll field.

Statistical Objective Analysis
example: SeaWiFS chlorophyll cont.

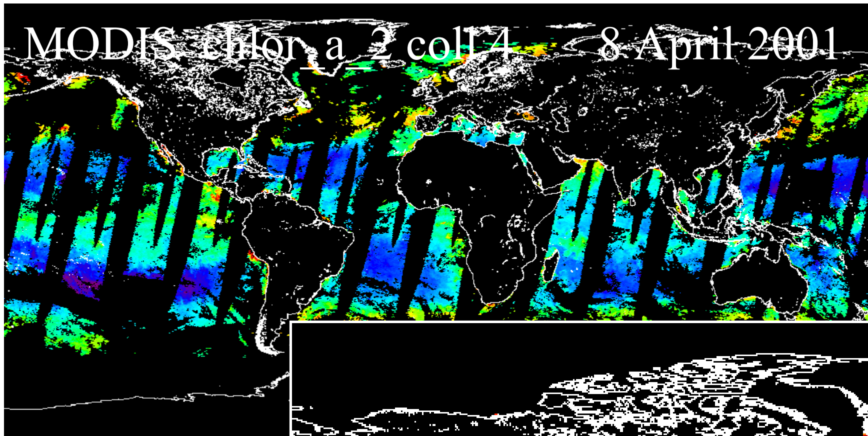
- Dependent on the chlorophyll increment space/time-lag correlations (i.e. an increment referenced to the background field).

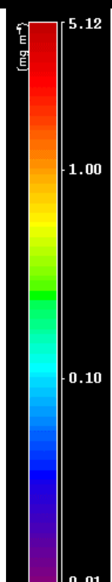


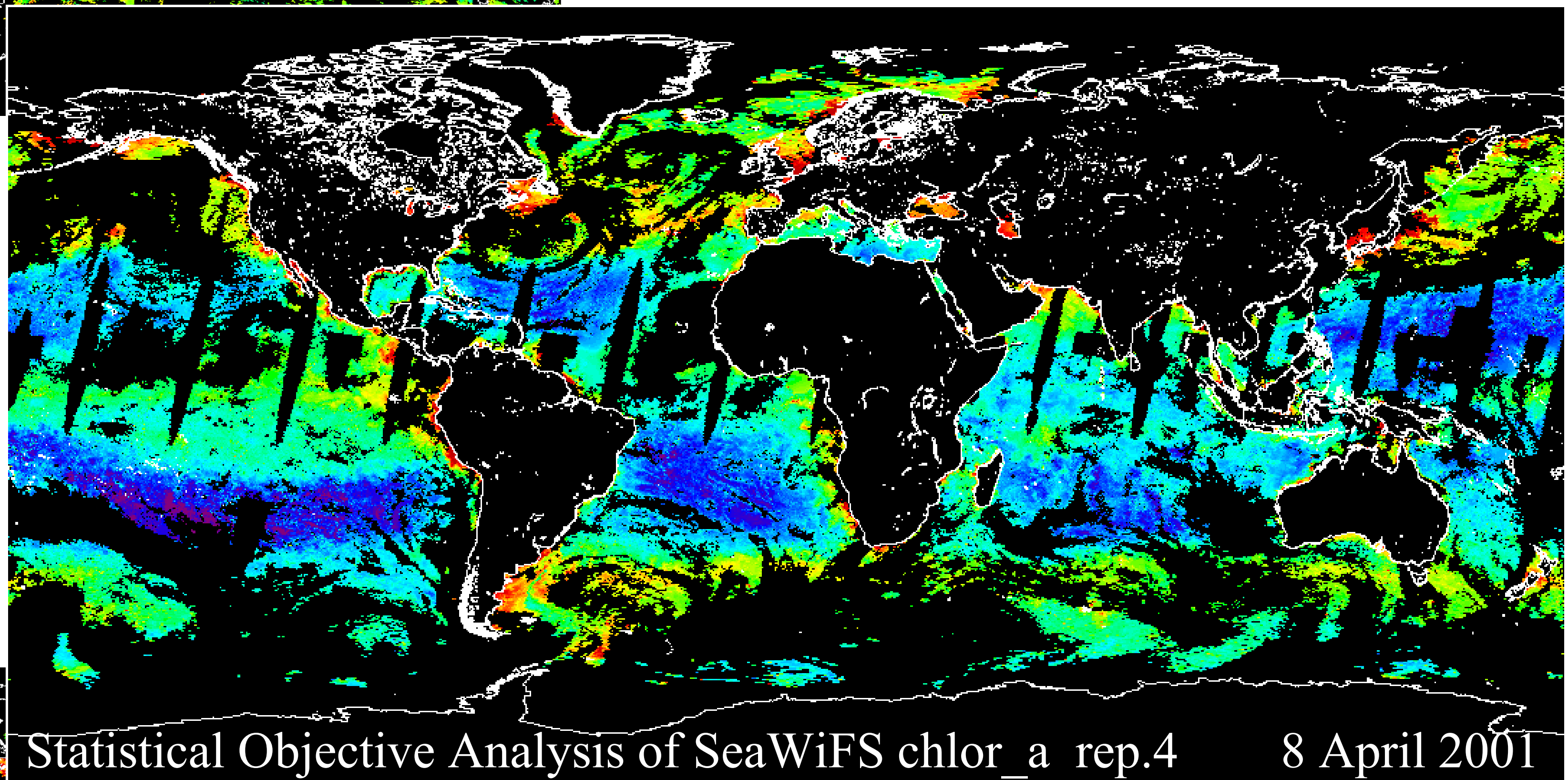


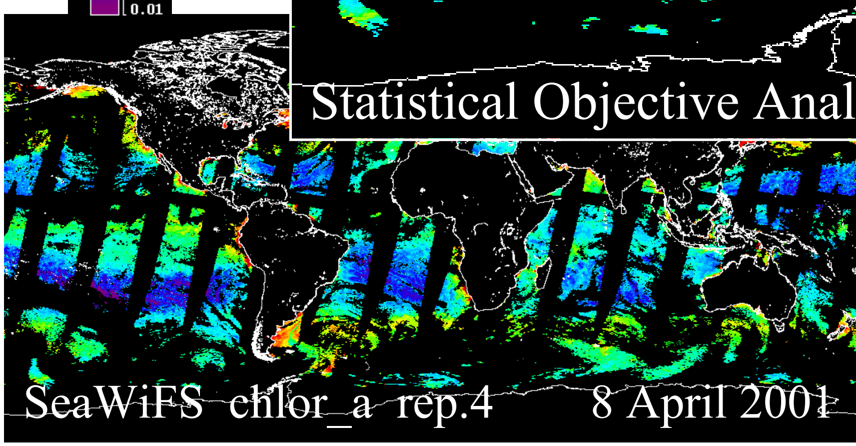
- Small radius of influence for the chlorophyll increment correlations (~350km).

2D Statistical Objective Analysis
(through space only)





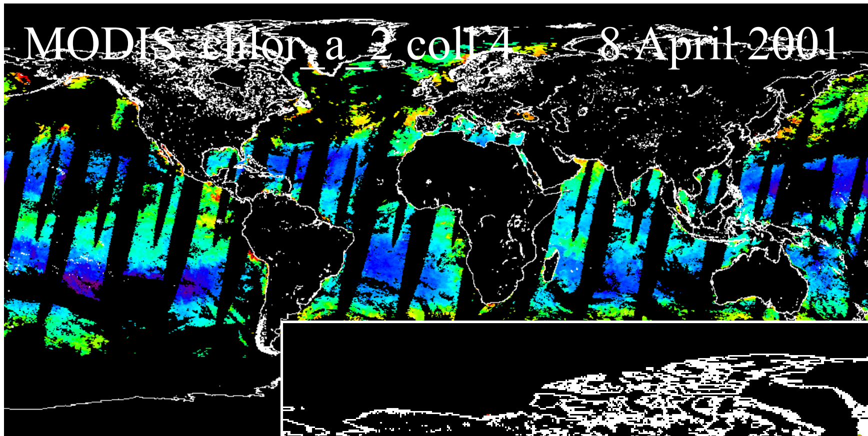


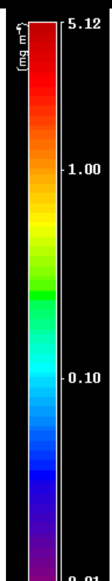


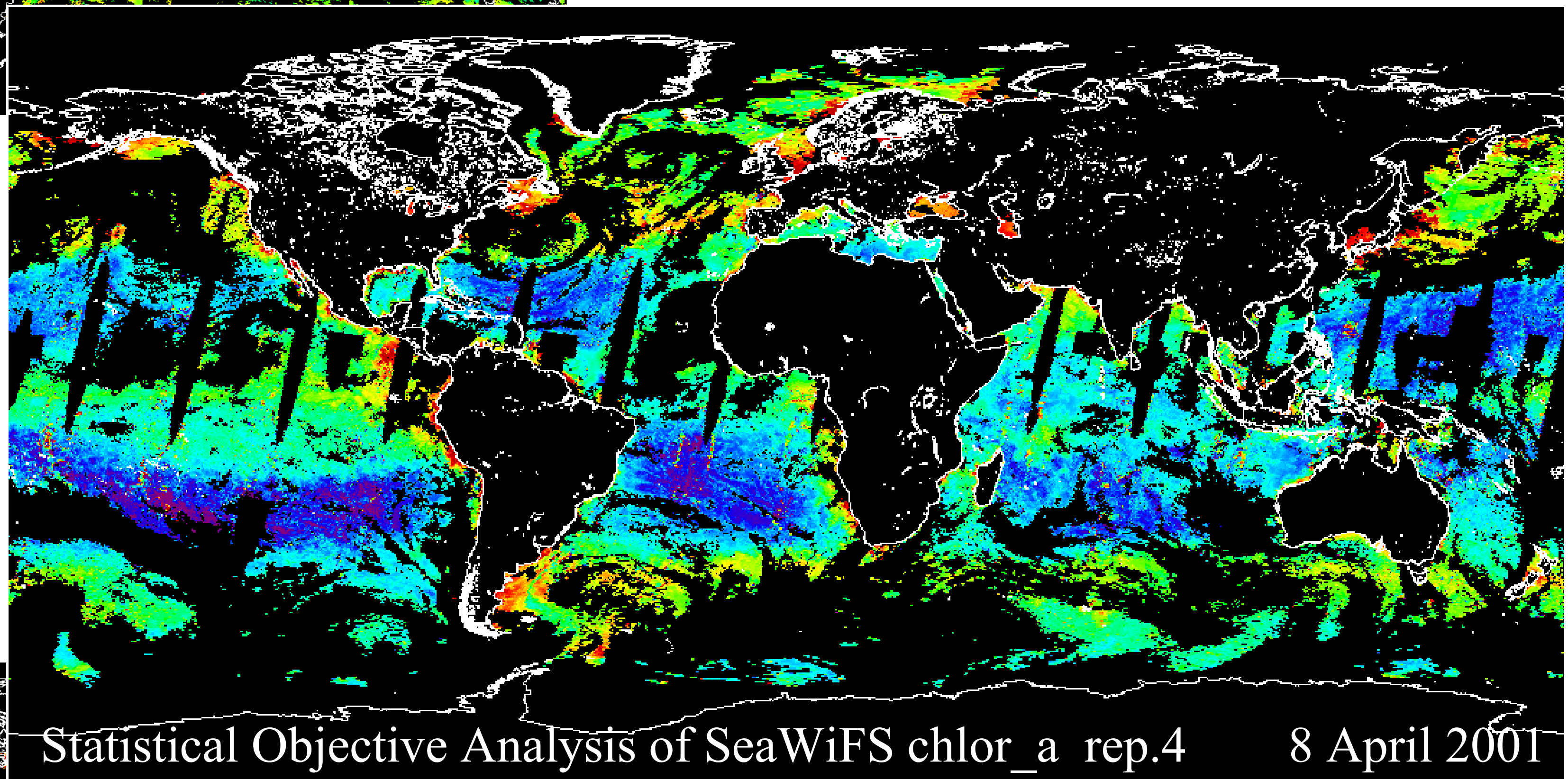
◆ Result of 2D statistical objective analysis on SeaWiFS data.

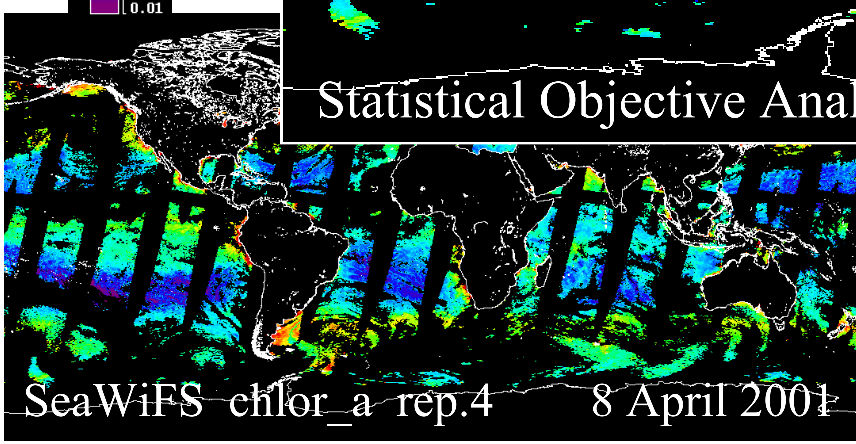
◆ Performed on the missing SeaWiFS points that coincide with the existing MODIS coverage.

3D Statistical Objective Analysis
(through space and time)









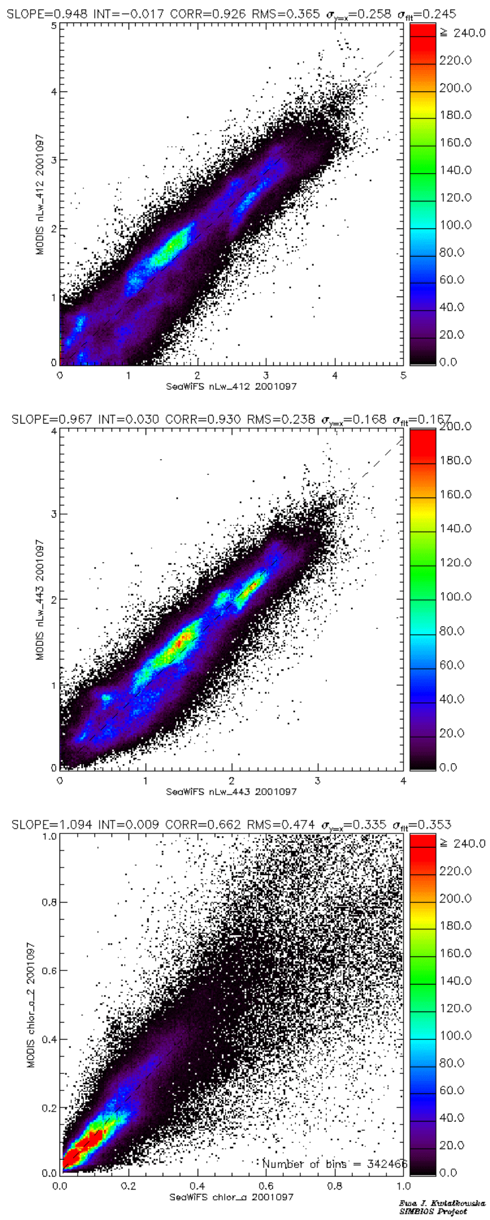
◆ Result of 3D statistical objective analysis on SeaWiFS data.

◆ Performed on the missing SeaWiFS points that coincide with the existing MODIS coverage.

Statistical Objective Analysis Notes

- It is difficult to establish reliable statistics for the analysis which is independent of point locations, i.e. the variance of chlorophyll data and the variance of chlorophyll noise.
- Large data gaps cannot be interpolated because of the limited radius of influence for the chlorophyll increment correlations (especially with 2D analysis).
- Low correlation values cause some interpolated points to be close to the background field value.
- In the 3D analysis, negative correlations across time cause undesirable effects to some interpolated data.
- Method is computationally demanding.

Regression Mapping



- Relationships between ocean color data from different sensors are complex and noisy.
- Ideally, the goal would be to bring multi-instrument, multi-platform, and multi-year ocean color observations to a ***common baseline***, consistently calibrated and jointly validated.
- Mapping of multi-sensor and multi-year ocean color data to a common baseline can be accomplished using ***regression***.

SIMBIOS
http://simbios.gsfc.nasa.gov

Neural-network/support-vector-machine Regression Mapping

- Regression mapping reproduces/emulates the response from the baseline given data from different sensors.
- NN/SVM – model-less regression based on machine learning from examples.
- Dependent on the choice of input feature elements, the architecture of the mapper, and training data.
- Regression is performed between individual data points, although spatial information can also be included.

SIMBIOS

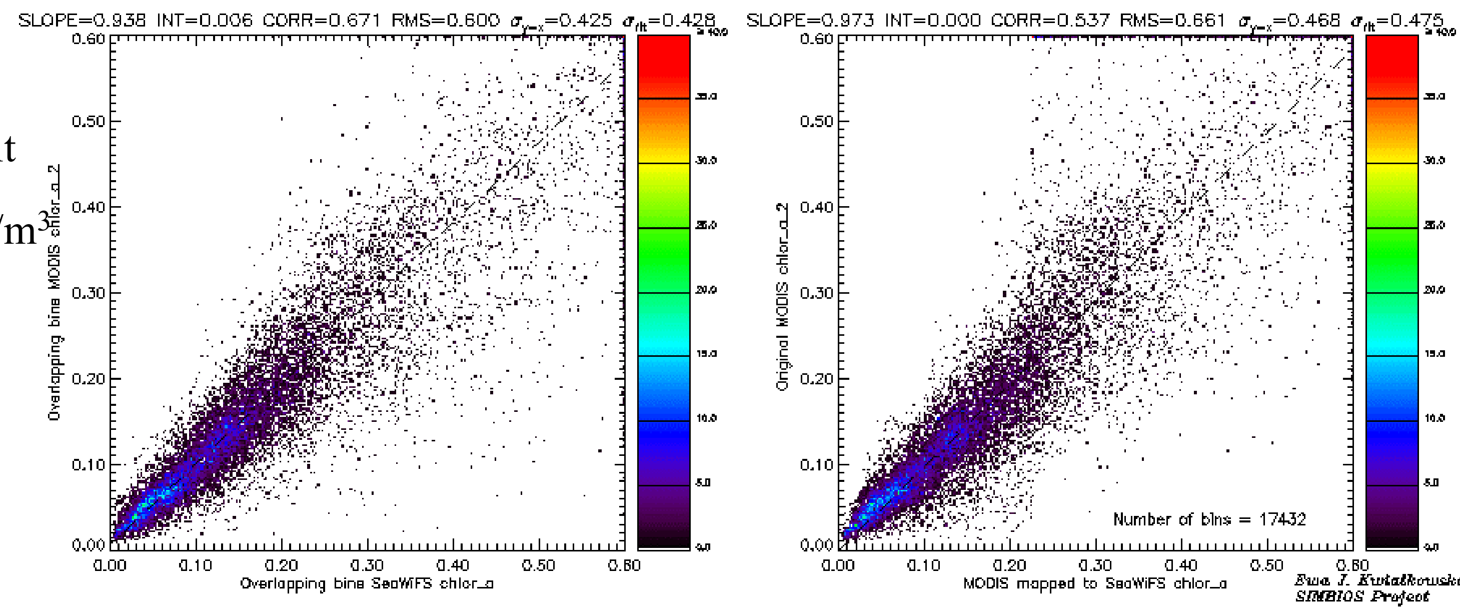
http://simbios.gsfc.nasa.gov

NN/SVM Regression Mapping

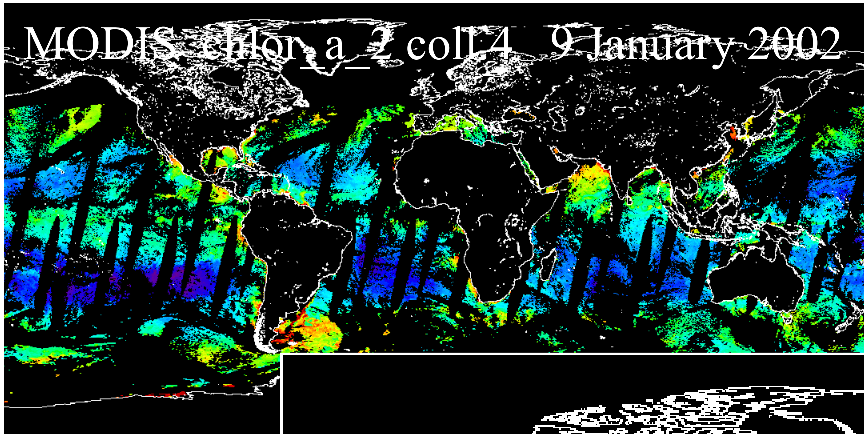
example: from MODIS data to SeaWiFS chlorophyll (SeaWiFS=baseline)

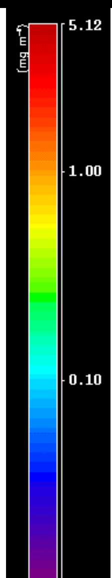
- MODIS input feature elements: nLw, chlorophyll, τ, ε, K490, senz, sunz, humidity, ozone, lat, lon, date.
- SeaWiFS baseline output: chlorophyll.
- Overlapping MODIS and SeaWiFS bins are divided into training and testing sets.

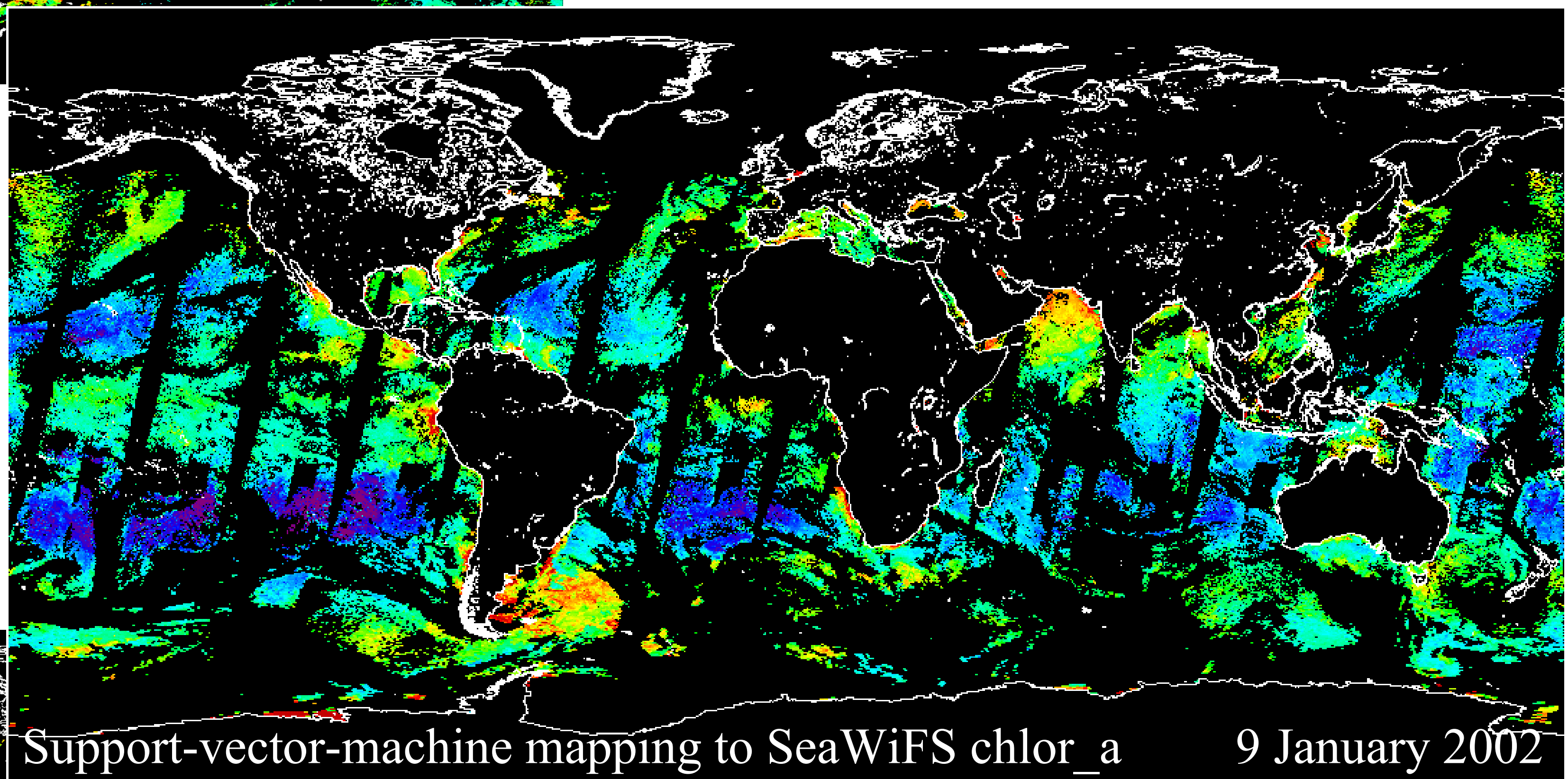
testing set result MAE = 0.09mg/m^3

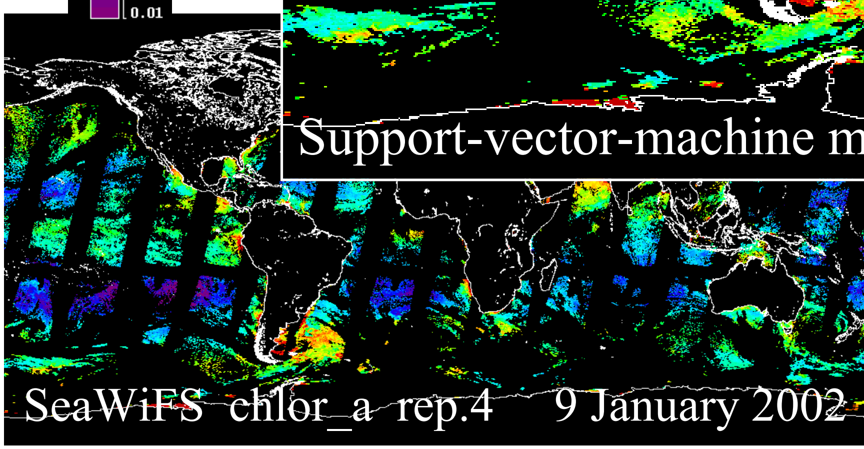


Support-vector-machine Regression Mapping



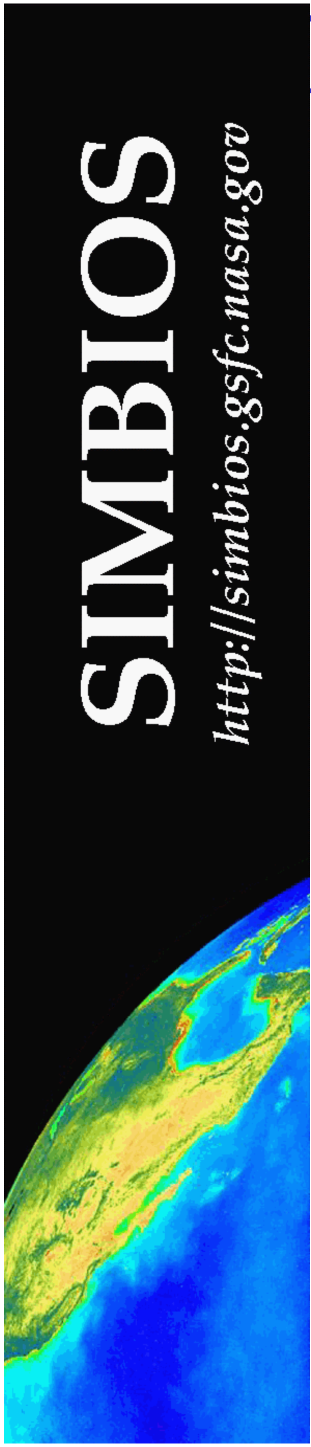






◆ Result of support-vector-machine mapping of unique MODIS coverage to SeaWiFS data.

◆ Global MODIS ∪ SeaWiFS coverage is represented in SeaWiFS chlorophyll values.



Neural-network/support-vector-machine Regression Mapping Notes

- Ocean color data pose a very complex regression problem, highly nonlinear and high multidimensional.
- Things to consider:
  - Choice of optimal input feature elements,
  - Comprehensive training sets encompassing a variety of ocean color, atmospheric, sensor viewing, and other conditions,
  - Preprocessing options for input feature elements, such as logarithm, scaling, and translation,
  - SVM kernels and NN topology and parameters.
- Once trained, the algorithm has minimal computational requirements.

Application of Spatial/Temporal Interpolation and Regression Mapping

- Data integration via interpolation and mapping does not have to be for mission ultimate products, such as chlorophyll, instead these can be
  - nLw,
  - radiances at the top of the atmosphere,
  - other sensor characteristics, e.g. contributing to instrument cross-calibration,
- Application within the spectral optimization algorithm of Chomko and Gordon.

Conclusions

- Data merger → create high quality long-term multi-sensor ocean color data sets.
- Bring all multi-sensor data to a common consistently calibrated and jointly validated ocean color baseline.
- Accomplish it through sensor intercomparisons, validations, cross-calibrations, and “intelligent” data fusion.
- Integrate data on three levels: the ultimate ocean color products (e.g. chlorophyll), the interim products (e.g. nLw’s), and the radiances at the top-of-the-atmosphere.